AD-A111 015 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/1
AN ALGORITHM FOR ESTIMATION IN SIMULTANEOUS LOGIT MODELS.(U)
DEC 81 M M MEYER N00014-80-C-0637
UNCLASSIFIED TR-226 NL

1 OF 1
AD
A111015

END
DATE
FILMED
3-82
DTIC

LEVEL II

AD A111015

DEPARTMENT

OF

STATISTICS

DTIC FILE COPY

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

82 02 17 063

LEVEL II

1

AN ALGORITHM FOR ESTIMATION IN SIMULTANEOUS LOGIT MODELS

by

Michael M. Meyer

Technical Report No. 226

Dept. of Statistics

Carnegie-Mellon University

Pittsburgh, PA 15213

December, 1981

This research was partially supported by the Office of Naval Research Contract N00014-80-c-0637 at Carnegie-Mellon University. I would like to S. E. Fienberg for helpful discussions during the preparation of this work.

TABLE OF CONTENTS

SUMMARY

This paper uses a very general version of the Iterative Proportional Fitting Procedure to develop an algorithm for estimation in simultaneous logit models. The algorithm can be used for any loglinear model which can be cast in the form of simultaneous logit equations. The principal advantage of this method is that it is not necessary to fit parameters associated with the sampling constraints and thus very large problems can be attacked. A numerical example using GLIM and a sample GLIM macro are included.

KEY WORDS: Categorical Data, Loglinear Models, Iterative Proportional Fitting Procedure.

1. INTRODUCTION

Simultaneous logit models have been oft-proposed but have suffered from a lack of a suitable and easy estimation technique. We show how an application of some simple ideas which underly the Iterative Proportional Fitting Procedure (IPFP) leads to a simple and worthwhile algorithm. The same approach can be used for any loglinear model and is potentially useful for large contingency table problems

Methods for analyzing data in the form of a binomial response variable with one or more covariates are well known. A common method, and the one which corresponds to a loglinear model, is logit regression. There are numerous programs (e.g., BMDPLR and GLIM) which will find maximum likelihood estimates for the logit regression model. When the response variable is multinomial, rather than binomial, the situation is rather different. The models which correspond to logit regression posit a system of simultaneous logit models (see, e.g., Fienberg (1980), Fienberg and Mason (1978) and Nerlove and Press (1973)). One method of finding M.L.E.'s in this case is to transform the simultaneous logit model into an equivalent loglinear model. This approach, however, introduces potentially many nuisance parameters corresponding to the "interactions" among the independent variables. It is possible that the introduction of these extra parameters will make the model too large to be fit conveniently. To overcome these problems Fienberg (1980) suggests fitting continuation ratios instead of the maximum likelihood analysis. We will show that this procedure is actually the first step in an IPFP and that the full IPFP is itself relatively easy to calculate. We also give a simple GLIM macro for this analysis.

The method is developed in several steps. First we consider a multinomial response variable with categorical explanatory variables. We then generalize this to continuous explanatory variables and finally consider models more complicated than those normally discussed.

In order to develop the method we need a brief discussion of some theory underlying the IPFP.

2. BACKGROUND AND NOTATION

Csiszar (1976) presents a very elegant and general discussion of the IPFP by developing a "geometry" for the information measure. A simplified version of the chief results of this theory are outlined below. Let n, p, q, r, s, and t denote p.m.f.'s which are non-zero for all elements of a finite set I. The Kullback-Leibler information number (or directed divergence) specifies a distance,

$$I(p\|q) = \frac{1}{|I|} \Sigma_{i \in I}\; p(i) \ln (p(i)/q(i))$$

between p and q. The principle of minimum discriminant information, as formulated by Kullback (1959), aims to minimize the distance between a reference distribution, q above, and a family of other distributions. The properties of such estimates have been studied extensively. The most important results can be found in Kullback (1959) and are summarized, with a special emphasis on contingency tables, in Gokhale and Kullback (1978). Darroch and Ratcliff (1972) also used the directed divergence in their development of the Generalized Iterative Scaling algorithm.

We need to specify an appropriate family, E, of probability mass functions (p.m.f.'s) over which to minimize the distance. Linear sets of p.m.f.'s are a natural class and correspond to usual loglinear models. A convex set, E, of p.m.f.'s is called linear if when p and q are in E and $t = \alpha \cdot p + (1-\alpha) \cdot q$ ($\alpha \in P$) is a p.m.f., then t is also in E. A p.m.f. which satisfies

$$I(q\|r) = \min_{p \in E} I\,(p\|r)$$

is called the I-projection of r on E and will be denoted by $q = P_E(r)$. Csiszar gives conditions under which $P_E(r)$ exists (it is always unique) and develops a geometry for I-projections by using an analogue of Pythagorous' Theorem. Now let $F = \{f_\gamma : \gamma \in \Gamma\}$ be a set of real valued functions on I and $A = \{a_\gamma : \gamma \in \Gamma\}$ be real constants. Define M_F to be span (F). A linear set, E, can be constructed by considering the set of p for which,

$$\sum_{i \in I} p(i) \cdot f_\gamma(i) = a_\gamma \;;\; \gamma \in \Gamma \,.$$

When we consider s to be an observed probability function and

$$a_\gamma = \sum_{i \in I} s(i)\, f_\gamma(i) \; : \; \gamma \in \Gamma$$

then the duality between maximum likelihood and minimum discriminant (see e.g. Darroch and Ratcliff (1972)) estimation states that if

$$\hat{q} = P_{\mathcal{E}}(r)$$

then

$$\ln(\hat{q}) \in M_{\mathcal{F}} + \ln(r)$$

and

$$\hat{q} - s \in M_{\mathcal{F}}^{\perp},$$

i.e. $\hat{q}$ is the m.l.e. (under Poisson sampling) for the corresponding log-affine model. Csiszar's principal theorem says that if $\mathcal{E}$ is the finite intersection of the linear sets $\mathcal{E}_k$ (i.e. $\mathcal{E} = \bigcup_{k \in K} \mathcal{E}_k$) then $\hat{q} = P_{\mathcal{E}}(r)$ is the pointwise limit of $q_n = P_{\mathcal{E}_n}(q_n)$ $n = 1,2,3 \ldots$ where $q_0 = r$ and $\mathcal{E}_n = \mathcal{E}_i$ if $i = n \bmod |K|$.

It is this theorem on cyclic projections that we shall use for the simultaneous logit algorithm. For more details about the above theory see Csiszar (1976) or Meyer (1981a) and for another application of the theory see Meyer (1981b). We now present a short example to illustrate the notation.

Example 1

Let p be an observed 3x3 probability function obtained via multinomial sampling and consider the model

$$\mathcal{E}(p(ij)) = q(i,j)\ ,$$

and $\ln(q(i,j)) = \mu + \alpha_i + \beta_j$; i,j = 1,2,3, i.e. independence of row and column categories.

The linear manifold for this model is spanned by a set of tables, f_R^i and f_C^j, where

$$f_R^i(k,\ell) = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}$$

and

$$f_C^j(k,\ell) = \begin{cases} 1 & \ell = j \\ 0 & \ell \neq j \end{cases}$$

Now let
$$F_R = \{ f_R^i ; i = 1,2,3 \} .$$
$$F_C = \{ f_c^j ; j = 1,2,3 \} .$$

and
$$F = F_R \cup F_c .$$

We further allow
$$a_R^i = \Sigma_{k,\ell}\ p(k,\ell)\ f_R^i(k,\ell).$$

and
$$a_C^j = \Sigma_{k,\ell}\ p(k,\ell)\ f_C^j(k,\ell)$$

and define E_R and E_C to be the linear spaces generated by the f_R, a_R and f_C, a_C pairs.

The maximum likelihood estimate $\hat{q}(i,j)$ can be generated by taking $\hat{q}_0(i,j) = 1\ \forall i,j$ and successively forming the I-projections onto E_R and E_C. This algorithm is just the usual IPFP and converges after one step.

3. SIMPLE SIMULTANEOUS LOGIT MODELS FOR A THREE-WAY TABLE

In this section we consider simultaneous logit models for three way tables where there is one response variable and two explanatory variables. In the next section we consider the general simultaneous logit case.

Consider a multinomial response variable A (with I levels) and explanatory variables B (with J levels) and C (with k levels). Denote the table of observed counts by

$$w_{ijk} \quad (i = 1,\dots, I;\ j = 1,\dots, J;\ k = 1,\dots, K)$$

where w_{ijk} is the number of responses at level i when the explanatory variables, B and C, are at levels j and k respectively. In this discussion we transform the data w into a probability vector, $z_{ijk} = w_{ijk} / w_{+++}$. This is not necessary in practice. We will denote the expected cell probability $E(z_{ijk})$ by m_{ijk} .

When I = 2 the saturated logit regression model, using the notation of Fienberg (1980), is

$$\text{logit}\ (m_{ijk}) = v + v_{2(j)} + v_{3(k)} + v_{23(jk)} .$$

If we leave out the interaction term, v_{23}, the model becomes

$$\text{logit}\ (m_{ijk}) = v + v_{2(j)} + v_{3(k)}$$

with likelihood equations

$$\hat{m}_{ij+} = z_{ij+}$$

$$\hat{m}_{i+k} = z_{i+k}$$

and the sampling constraint

$$\hat{m}_{+jk} = z_{+jk}\ .$$

It is a simple matter to rewrite this model as a loglinear model, viz., $\ln(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$ with likelihood equations

$$\hat{m}_{ij+} = z_{ij+}$$

$$\hat{m}_{i+k} = z_{i+k}$$

$$\hat{m}_{+jk} = z_{+jk}\ .$$

The logit model was defined only for $I = 2$ but the loglinear model is defined for any value of I. If we consider this arbitrary loglinear model we can write it as the set of simultaneous logit models

(3.1–1) $$\ln(\hat{m}_{1jk}/\hat{m}_{2jk}) = v^1 + v^1_{1(j)} + v^2_{2(k)}$$

(3.1–2) $$\ln(\hat{m}_{2jk}/\hat{m}_{3jk}) = v^2 + v^2_{1(j)} + v^2_{2(k)}$$

$$\vdots$$

(3.1–(I–1)) $$\ln(\hat{m}_{I-1,j,k}/\hat{m}_{Ijk}) = v^{I-1} + v^{I-1}_{1(j)} + v^{I-1}_{2(k)}$$

(3.1–(I)) $$\ln(\hat{m}_{Ijk}/\hat{m}_{1jk}) = v^I + v^I_{1(j)} + v^I_{2(k)}$$

The last equation, (3.1–(I)) is redundant but it doesn't hurt to include it and it may actually improve the resulting algorithm. We shall similarly break up the likelihood equations into

$$\hat{m}_{1j+} = z_{1j+}\ ,\ \hat{m}_{2j+} = z_{2j+}$$

(3.2–1) $$\hat{m}_{1+k} = z_{1+k}\ ,\ \hat{m}_{2+k} = z_{2+k}$$

$$\hat{m}_{1jk} + \hat{m}_{2jk} = z_{1jk} + z_{2jk}$$

$$\hat{m}_{2j+} = z_{2j+},\ \hat{m}_{3j+} = z_{3j+}$$

(3.2-2) $$\hat{m}_{2+k} = z_{2+k},\ \hat{m}_{3+k} = z_{3+k}$$

$$\hat{m}_{2jk} + \hat{m}_{3jk} = z_{2jk} + z_{3jk}$$

$$\vdots$$

$$\hat{m}_{I-1,j+} = z_{I-1,j+},\ m_{Ij+} = z_{Ij+}$$

(3.2-(I-1)) $$\hat{m}_{I-1,+k} = z_{I-1,+k},\ \hat{m}_{I+k} = z_{I+k}$$

$$\hat{m}_{I-1,jk} + \hat{m}_{Ijk} = z_{I-1,jk} + z_{Ijk}$$

$$\hat{m}_{Ij+} = z_{Ij+},\ \hat{m}_{1j+} = z_{1j+}$$

(3.2-(I)) $$\hat{m}_{I+k} = z_{I+k},\ \hat{m}_{1+k} = z_{1+k}$$

$$\hat{m}_{Ijk} + \hat{m}_{1jk} = z_{Ijk} + z_{1jk}$$

To fit the full loglinear model would require IJK $-$ (I$-$1)(J$-$1)(K$-$1) parameters, perhaps a very large number. Each of the logit models requires only J + K $-$ 1 parameters. If we needed to fit only the logit models then the problem would be easy.

Consider the following algorithm. Let $\hat{m}^0_{ijk} = 1,\ \forall$ ijk. Estimate $\hat{m}^1_{1jk}$ and $\hat{m}^1_{2jk}$ using the logit model (3.1-1) with likelihood equations (3.2-1) and set $\hat{m}^1_{ijk} = \hat{m}^0_{ijk}$ for $i \neq 1,2$. Now estimate $\hat{m}^2_{2jk}$ and $\hat{m}^2_{3jk}$ from (3.1-2) and (3.2-2) with $\hat{m}^2_{ijk} = \hat{m}^1_{ijk}$ for $i \neq 2,3$. Continue cycling through the logit equations, each time adjusting one pair of the multinomials leaving the rest untouched. Each iteration of the algorithm is a (small) logit regression problem with an entire cycle requiring I logit regressions. We will now show that $\hat{m}^{\ell}_{ijk}$ converges to the M.L.E. $\hat{m}_{ijk}$.

If we view the problem from the I-divergence point of view we know that we require

$$\hat{m} = P_{\mathcal{E}}(\hat{m}^0)$$

where
$$\mathcal{E} = \left\{ \text{p.m.f.'s } p_{ijk} : \begin{array}{l} p_{ij+} = z_{ij+} \\ p_{i+k} = z_{i+k} \\ p_{+jk} = z_{+jk} \end{array} \right\}$$

If we view the logit regressions as merely calculating I-projections we note that

$$\hat{m}^1 = P_{E_1}(\hat{m}^0)$$

where

$$E_1 = \{\text{p.m.f.'s } p_{ijk} \text{ satisfying (3.2-1)}\}$$

and

$$\hat{m}^2 = P_{E_2}(\hat{m}^1)$$

where

$$E_2 = \{\text{p.m.f.'s } p_{ijk} \text{ satisfying (3.2-2)}\}$$

etc.

As $\bigcap_{i=1}^{I} E_i = E$ (in fact $\bigcap_{i=1}^{I-1} E_i = E$) we can invoke the theorems on I-projections and state that $\hat{m}^{\ell}_{ijk}$ must converge to the M.L.E.

Thus the procedure of sequentially solving the logit equations is really a version of the IPFP. There is nothing special about the particular logit formulation we have used. All that is required is that the E_i corresponding to the logit models must have an intersection which is E. We note that some of the orderings suggested by Fienberg (1980, p.110) do not satisfy this requirement, although they are sensible for other reasons.

Generally it is not possible to obtain estimates of asymptotic (co)variance matrices using an IPFP algorithm. In this case, however, if the method used to solve the individual logit problem will yield asymptotic (co)variance estimates then this information can be used to compute (co)variance estimates for the multinomial logit problem. For example, if estimates for the v_i^{ℓ} and their covariance matrices for $\ell = 1, \ldots, I$ are obtained on the last cycle of the algorithm, these will be correct for the full problem. It is, however, impossible to find estimates of the correlations between, say, v_i^1 and v_i^3, for exactly the same reasons that covariances are unavailable in the usual IPFP.

In order to clarify the above discussion we present a small numerical example of a trinomial logit regression problem.

Example 2

Consider two explanatory variables, sex and intelligence, each with two levels. For each combination of levels we observe a trinomial response. The data are

		Response D	Response E	Response F
I(1)	S(1)	7	7	7
	S(2)	8	24	20
I(2)	S(1)	9	6	16
	S(2)	26	11	16

The simultaneous logit model we consider is

$$\ln(D/E) = G^1 + I^1_{(i)} + S^1_{(j)}$$

$$\ln(E/F) = G^2 + I^2_{(i)} + S^2_{(j)}$$

and

$$\ln(F/D) = G^3 + I^3_{(i)} + S^3_{(j)}$$

Each of these logit models requires 3 parameters whereas the equivalent loglinear model has 10 parameters (6 corresponding to two logit models and 4 from the sampling constraints). By using a loglinear model algorithm we obtain the M.L.E. for the cell means as approximately

		Response D	Response E	Response F
I(1)	S(1)	4.035	7.634	9.331
	S(2)	10.965	23.366	17.668
I(2)	S(1)	11.965	5.366	13.668
	S(2)	23.035	11.634	18.331

After one cycle of the simultaneous logit approach (i.e., 3 logit models) the fitted values are

		Response D	Response E	Response F
I(1)	S(1)	4.217	7.163	9.621
	S(2)	10.783	23.837	17.379
I(2)	S(1)	11.783	5.837	13.379
	S(2)	23.217	11.163	18.621

and after [illegible] cycles the fitted values are the same as those from the loglinear model algorithm.

If we constrain the parameters so that $I^{\ell}_{(1)}$ and $S^{\ell}_{(1)}$ are zero, for ℓ = 1,2,3 , then the estimated parameter values and their standard errors are:

Parameters	Estimated parameter	Estimated s.e.
G^1	−.638	.4706
$I^1_{(2)}$	−.119	.4737
$S^1_{(2)}$	1.440	.4327
G^2	−.201	.4015
$I^2_{(2)}$	.480	.4278
$S^2_{(2)}$	−.734	.4023
G^3	.839	.4341
$I^3_{(2)}$	−.361	.4111
$S^3_{(2)}$	−.705	.4074

While it is possible to calculate the estimated covariance between, say, $\hat{G}^1$ and $\hat{I}^1_{(2)}$ from the Newton algorithm used for the logit regressions, it is not possible to estimate the covariance between, say, $\hat{G}^1$ and $\hat{G}^2$.

The calculations for this example were performed using GLIM and the complete output and macros are in the Appendix.

4. GENERAL SIMULTANEOUS LOGIT MODELS

Thus far we have only touched the surface of models which can be fit using this technique, but the general form is the same as the specific example we have considered. Let $\mathcal{J}$ = {1,2,..., J} be an index set and suppose that for each element of $\mathcal{J}$ we observe an I-nomial response. Denote the response by x_{ij}, i = 1,..., I, $j \in \mathcal{J}$ and the expected count by m_{ij}. In other words we have product-multinomial sampling with x_{+j} observations in category j. Any of the usual factorial loglinear models which condition on the explanatory variables can be written in the simultaneous logit form,

$$\ln(m_1/m_2) \in M$$
$$\ln(m_2/m_3) \in M$$
$$\vdots$$
$$\ln(m_I/m_1) \in M$$

where M is a linear manifold in R^J. The linear manifold M determines a linear space of p.m.f.'s E. Just as in the previous case, cycling through the logit equations leads to the M.L.E.'s for the corresponding loglinear model.

In the above formulation there is no need for each logit to belong to the same linear manifold. In other words, the most general form of loglinear model which satisfies the sampling constraints can be written as

$$\ln(m_1/m_2) \in M_1$$
$$\ln(m_2/m_3) \in M_2$$
$$\vdots$$
$$\ln(m_{I-1}/m_I) \in M_{I-1}$$

where each $M_i \subset R^J$. Each M_i determines a linear space of P.D.'s, E_i a subset of the p.m.f.'s on I × J points, and sequentially fitting the logit models produces the I-projection onto $E = \bigcap_{i=1}^{I-1} E_i$. Thus any loglinear model which involves product-multinomial constraints, or can be written in this form, can be fit as a sequence of logit models. In this way it is possible to fit the loglinear model without necessarily fitting "parameters" for the product-multinomial constraints.

REFERENCES

Csiszar, I. (1975). I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability, 3,* 146-158.

Darroch, J. N. and Ratcliff, D. (1972). Generalized Iterative Scaling for Log-Linear Models. *The Annals of Mathematical Statistics, 11,* 427-444.

Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data.* Cambridge, MA: The MIT Press.

Fienberg, S .E. and Mason, W. M. (1978). Identification and Estimation of Age-period-cohort Effects in the Analysis of Discrete Archival Data. *Sociological Methodology 1979,* 1-67.

Gokhale, D. V. and Kullback, S. (1978). *The Information in Contingency Tables.* New York: Marcel Dekker.

Kullback, S. (1959). *Information Theory and Statistics.* New York: John Wiley & Sons.

Meyer, M. M. (1981a). *Applications and Generalizations of the Iterative Proportional Fitting Procedure.* Ph.D. Thesis, School of Statistics, University of Minnesota.

Meyer, M. M. (1981b). *Transforming Contingency Tables.* Technical Report 214, Department of Statistics, Carnegie-Mellon University.

Nerlove, M. and Press, S. J. (1973). *Univariate and Multivariate Log-linear and Logistic Models.* Rand Corp. Technical Report R-1306-EDA/NIH. Santa Monical CA.

APPENDIX I. A GLIM MACRO AND AN EXAMPLE

I.1. GLIM MACROS

The following GLIM macros were used to calculate the multinomial logit estimates of Section 3.

```
$SUB MLOG
$PRI '*DELETE logithelp  IF NOT NEEDED * '
$MAC LOGITHELP !
 There are 6 macros, 3 which include parameter estimates and three which don't.!
   WOUT - Basic macro with parameter estimates, requires macros  DM  and  PARA.!
   NOUT - Basic macro without estimates,  requires macros  DD  and LOG.!
!
 Typically one would use   NOUT  until satisfied that the process has converged!
 and then  WOUT  , once, to get parameter estimates.!
 NOUT  and  WOUT  have up to 9 arguments, one for each response level.!
 Initialize the   n with  $ARG NOUT A B ... I , where A B ... I are identifiers!
 for the response levels.  The scalar  %N  must be set to the number of !
 response levels and  NOUT  and  WOUT  must have %N arguments.!
 To get started, one needs to fit the model to some data, any will do.!
 The identifiers  A B ... I are overwritten with the fitted values. !
 so copy them or you'll loose them.!
$ENDM


$MAC NOUT !
$CAL %D=0 !
$OUT $ARG LOG %1 %2 %3 %4 %5 %6 %7 %8 %9 $USE LOG !
$USE DD $OUT 5 $PRI 'CONVERGENCE    ' %D $END
$MAC DD !
$YVAR %1 $CAL ZZ=%1+%2 $ERR B ZZ $FIT . $CAL %D=%IF(%GT(%DV,%D),%DV,%D)!
$CAL %1=%FV $CAL %2=ZZ-%1 $END
$MAC LOG !
$ARG DD %1 %2 $USE DD $ARG DD %2 %1 $CAL %M=%EQ(%N,2) $EXIT %M !
$ARG DD %2 %3 $USE DD $ARG DD %3 %1 $CAL %M=%EQ(%N,3) $EXIT %M !
$ARG DD %3 %4 $USE DD $ARG DD %4 %1 $CAL %M=%EQ(%N,4) $EXIT %M !
$ARG DD %4 %5 $USE DD $ARG DD %5 %1 $CAL %M=%EQ(%N,5) $EXIT %M !
$ARG DD %5 %6 $USE DD $ARG DD %6 %1 $CAL %M=%EQ(%N,6) $EXIT %M !
$ARG DD %6 %7 $USE DD $ARG DD %7 %1 $CAL %M=%EQ(%N,7) $EXIT %M !
$ARG DD %7 %8 $USE DD $ARG DD %8 %1 $CAL %M=%EQ(%N,8) $EXIT %M !
$ARG DD %8 %9 $USE DD $ARG DD %9 %1 $END

$MAC WOUT !
$CAL %D=0 !
$ARG PARA %1 %2 %3 %4 %5 %6 %7 %8 %9 $USE PARA !
$USE DM $PRI 'CONVERGENCE    ' %D $END
$C  need to get routines for goodness of fit here.
$MAC PARA !
$ARG DM %1 %2 $USE DM $ARG DM %2 %1 $CAL %M=%EQ(%N,2) $EXIT %M !
```

```
$ARG DM %2 %3 $USE DM $ARG DM %3 %1 $CAL %M=%EQ(%N,3) $EXIT %M !
$ARG DM %3 %4 $USE DM $ARG DM %4 %1 $CAL %M=%EQ(%N,4) $EXIT %M !
$ARG DM %4 %5 $USE DM $ARG DM %5 %1 $CAL %M=%EQ(%N,5) $EXIT %M !
$ARG DM %5 %6 $USE DM $ARG DM %6 %1 $CAL %M=%EQ(%N,6) $EXIT %M !
$ARG DM %6 %7 $USE DM $ARG DM %7 %1 $CAL %M=%EQ(%N,7) $EXIT %M !
$ARG DM %7 %8 $USE DM $ARG DM %8 %1 $CAL %M=%EQ(%N,8) $EXIT %M !
$ARG DM %8 %9 $USE DM $ARG DM %9 %1 $END
$MAC DM !
$YVAR %1 $CAL ZZ=%1+%2 $ERR B ZZ $FIT . $CAL %D=%IF(%GT(%DV,%D),%DV,%D)!
$DIS E $CAL %1=%FV $CAL %2=ZZ-%1 $END
$RETURN
$FINISH
```

I.2. AN EXAMPLE USING THE MACROS

The complete GLIM output for Example 2 is presented below.

```
GLIM 3.09 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

?$INPUT 35 MLOG     $C ... INPUT THE MACROS  $

*DELETE logithelp  IF NOT NEEDED *
?$PRINT LOGITHELP $


 There are 6 macros, 3 which include parameter estimates and three which don't.
   WOUT - Basic macro with parameter estimates, requires macros  DM  and  PARA.
   NOUT - Basic macro without estimates,  requires macros  DD  and LOG.

 Typically one would use   NOUT  until satisfied that the process has converged
 and then  WOUT  , once, to get parameter estimates.
 NOUT  and  WOUT  have up to 9 arguments, one for each response level.
 Initialize the run with  $ARG NOUT A B ... I , where A B ... I are identifiers
 for the response levels.  The scalar  %N  must be set to the number of
 response levels and  NOUT  and  WOUT  must have %N arguments.
 To get started, one needs to fit the model to some data, any will do.
 The identifiers  A B ... I are overwritten with the fitted values,
 so copy them or you'll loose them.

?$DEL LOGITHELP $UNIT 4 $DATA D E F $READ

? 7 7 7   8 24 20   9 6 16   26 11 16 $

?$FACTOR I 2 S 2 $CAL I=%GL(2,2) :S=%GL(2,1) $

?$CAL T1=1 :T2=2 $YVAR T1 $ERR B T2 $C ... GET THE ALGORITHM STARTED $

?$FIT S+I $

        SCALED
```

```
 CYCLE  DEVIANCE      DF
    4   0.8195E-15     1

?$CAL A=D :B=E :C=F $C ... COPY THE DATA SO THAT IT ISN'T LOST $

?$CAL %N=3 $ARG NOUT A B C $USE NOUT $

CONVERGENCE       2.791
?$LOOK A B C $

   1     4.217       7.163       9.621
   2     10.78       23.84       17.38
   3     11.78       5.837       13.38
   4     23.22       11.16       18.62
?$C ... THESE ARE THE ESTIMATES AFTER THE FIRST ITERATION $

?$USE NOUT $

CONVERGENCE      0.0950

?$USE NOUT $

CONVERGENCE      0.0013
?$USE NOUT $

CONVERGENCE      0.0000
?

?$C ... NOW USE WOUT TO GET PARAMETER ESTIMATES. ETC. $

?$ARG WOUT A B C $USE WOUT $

        SCALED
 CYCLE  DEVIANCE      DF
    3   0.2651E-06     1

       ESTIMATE       S.E.      PARAMETER
   1 -0.6377        0.4705      %GM
   2 -0.1189        0.4737      S(2)
   3   1.440        0.4327      I(2)
  SCALE PARAMETER TAKEN AS     1.000

----- CURRENT DISPLAY INHIBITED
        SCALED
 CYCLE  DEVIANCE      DF
    3   0.8541E-07     1

       ESTIMATE       S.E.      PARAMETER
   1 -0.2008        0.4015      %GM
   2  0.4803        0.4278      S(2)
   3 -0.7342        0.4023      I(2)
  SCALE PARAMETER TAKEN AS     1.000
```

```
----- CURRENT DISPLAY INHIBITED
       SCALED
 CYCLE  DEVIANCE        DF
    3    0.3044E-07       1

        ESTIMATE        S.E.         PARAMETER
   1   0.8385          0.4341         %GM
   2  -0.3614          0.4111         S(2)
   3  -0.7054          0.4074         I(2)
  SCALE PARAMETER TAKEN AS      1.000

----- CURRENT DISPLAY INHIBITED
CONVERGENCE         0.0000
?$LOOK  A  B  C  $

   1      4.035          7.634          9.332
   2      10.97          23.37          17.67
   3      11.97          5.366          13.67
   4      23.03          11.63          18.33
?

?$C ... ONE COULD NOW WRITE MACROS TO TEST GOODNESS-OF-FIT OF
        A  B  C  WITH  D  E  F  $

?$STOP
```

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #226	2. GOVT ACCESSION NO. AD-A111015	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN ALGORITHM FOR ESTIMATION IN SIMULTANEOUS LOGIT MODELS		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael M. Meyer		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0637
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Carnegie-Mellon University Pittsburgh, PA 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Contracts Office Carnegie-Mellon University Pittsburgh, PA 15213		12. REPORT DATE 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Categorical Data, Loglinear Models, Iterative Proportional Fitting Procedure

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper uses a very general version of the Iterative Proportional Fitting Procedure to develop an algorithm for estimation in simultaneous logit models The algorithm can be used for any loglinear model which can be cast in the form of simultaneous logit equations. The principal advantage of this method is that it is not necessary to fit parameters associated with the sampling constraints and thus very large problems can be attacked. A numerical example using GLIM and a sample GLIM macro are included.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FILMED

3-8